PCT WORLD INTELLECTUAL PROPERTY ORGANIZATION International Bureau

INTERNATIONAL APPLICATION PUBLISHED UNDER THE PATENT COOPERATION TREATY (PCT)

(51) International Patent Classification 5 : C12N 15/12, C07K 7/10, C12N 5/10, A61K 37/02

A1

(11) International Publication Number: WO 94/14959

(43) International Publication Date: 7 July 1994 (07.07.94)

(21) International Application Number: PCT/EP93/03645

(22) International Filing Date: 21 December 1993 (21.12.93)

(30) Priority Data:
RM92A000919 22 December 1992 (22.12.92) IT

(71) Applicant *(for all designated States except US)*: APPLIED RESEARCH SYSTEMS ARS HOLDING N.V. [NL/NL]; 6 John Gorsiraweg, Curaçao (AN).

(72) Inventor; and

(75) Inventor/Applicant *(for US only)*: SIRNA, Antonino [IT/IT]; Via Monti di San Paolo, 126, I-00126 Rome (IT).

(74) Agent: GERVASI, Gemma; Notarbartolo & Gervasi S.r.l., Viale Bianca Maria, 33, I-20122 Milan (IT).

(81) Designated States: AU, BY, CA, FI, JP, KR, KZ, NO, RU, UA, US, European patent (AT, BE, CH, DE, DK, ES, FR, GB, GR, IE, IT, LU, MC, NL, PT, SE).

Published

With international search report.

Before the expiration of the time limit for amending the claims and to be republished in the event of the receipt of amendments.

(54) Title: NEW PROTEIN FROM URINE NAMED COMPONENT B

(57) Abstract

A new protein is described obtainable from urine through an extraction and purification process by ion-exchange chromatography and high resolution chromatography, with a molecular weight of about gkDa and having anti-inflammatory, anti-coagulant and anti-tumoral properties. The protein inhibits binding of TGF-alpha to its receptor.

FOR THE PURPOSES OF INFORMATION ONLY

Codes used to identify States party to the PCT on the front pages of pamphlets publishing international applications under the PCT.

AT	Austria	**GB**	United Kingdom	**MR**	Mauritania
AU	Australia	**GE**	Georgia	**MW**	Malawi
BB	Barbados	**GN**	Guinea	**NE**	Niger
BE	Belgium	**GR**	Greece	**NL**	Netherlands
BF	Burkina Faso	**HU**	Hungary	**NO**	Norway
BG	Bulgaria	**IE**	Ireland	**NZ**	New Zealand
BJ	Benin	**IT**	Italy	**PL**	Poland
BR	Brazil	**JP**	Japan	**PT**	Portugal
BY	Belarus	**KE**	Kenya	**RO**	Romania
CA	Canada	**KG**	Kyrgystan	**RU**	Russian Federation
CF	Central African Republic	**KP**	Democratic People's Republic of Korea	**SD**	Sudan
CG	Congo			**SE**	Sweden
CH	Switzerland	**KR**	Republic of Korea	**SI**	Slovenia
CI	Côte d'Ivoire	**KZ**	Kazakhstan	**SK**	Slovakia
CM	Cameroon	**LI**	Liechtenstein	**SN**	Senegal
CN	China	**LK**	Sri Lanka	**TD**	Chad
CS	Czechoslovakia	**LU**	Luxembourg	**TG**	Togo
CZ	Czech Republic	**LV**	Latvia	**TJ**	Tajikistan
DE	Germany	**MC**	Monaco	**TT**	Trinidad and Tobago
DK	Denmark	**MD**	Republic of Moldova	**UA**	Ukraine
ES	Spain	**MG**	Madagascar	**US**	United States of America
FI	Finland	**ML**	Mali	**UZ**	Uzbekistan
FR	France	**MN**	Mongolia	**VN**	Viet Nam
GA	Gabon				

NEW PROTEIN FROM URINE NAMED COMPONENT B

The present invention relates to a new protein named Component B. In particular the invention relates to a new protein obtainable from urine, its preparation from urine, its production by recombinant DNA techniques using genomic DNA or cDNA encoding said new protein, as well as pharmaceutical compositions containing it and its use in therapy.

A new protein was isolated during the extraction and purification process of urine derivatives, this protein shows a polypeptide nature and relatively low molecular weight. When humane urine is treated with adsorbing materials, as kaolin, and then undergoes filtration, ion exchange chromatography and high resolution chromatography, preferably according to the process hereafter described, after lyophilisation a compound is obtained as amorphous white powder, moving as a single peak in high pressure reversed phase liquid chromatography (HPLC-RP) and having a molecular weight of about 9KDa when analysed by electrophoresis on a polyacrylamide gel in the presence of sodium dodecyl sulphate (SDS-Page) under reducing conditions. This protein was named, and is referred to hereinafter, Component B.

Component B is more specifically characterised through the amino acid sequence reported as SEQ ID NO: 1.

The present invention makes therefore available a new protein, named Component B, obtainable through a process comprising the isolation of a raw fraction of the compound itself from a dialysed concentrate

of urine after treatment with an adsorbing agent and its purification by ion exchange chromatography and high resolution chromatography as described hereafter.

Preferably the protein according to the present invention is extracted from human urine because of its high amount available useful for industrial production. The present invention refers particularly to a polypeptide comprising the SEQ ID NO: 1, its salts, functional derivatives, precursors and active fractions as well as its active mutants, i.e. other proteins or polypeptides wherein one or more amino acids of the structure were eliminated or substituted by other amino acids or one or more amino acids were added to that sequence in order to obtain polypeptides or proteins having the same activity of Component B and comprises also the corresponding fusion proteins i.e. polypeptides comprising Component B or a mutation thereof fused with another protein and having a longer lasting half-life in body fluids. Component B can therefore be fused with another protein such as, for example, an immunoglobulin.

The definition "salts" as used herein refers both to salts of the carboxyl-groups and to the salts of the amino functions of the compound obtainable through known methods.

The salts of the carboxyl-groups comprise inorganic salts as, for example, sodium, potassium, calcium salts and salts with organic bases as those formed with an amine as triethanolamine, arginine or lisine. The salts of the amino groups comprise for example salts with inorganic acids as hydrochloric acid and with organic acids as

acetic acid. The definition "functional derivatives" as herein used refers to derivatives which can be prepared from the functional groups present on the lateral chains of the amino acid moieties or on the terminal N- or C- groups according to known methods and are comprised in the invention when they are pharmacetically acceptable i.e. when they do not destroy the protein activity or do not impart toxicity to the pharmaceutical compositions containing them.

Such derivatives include for example esters or aliphatic amides of the carboxyl-groups and N-acyl derivatives of free amino groups or O-acyl derivatives of free hydroxyl-groups and are formed with acyl-groups as for example alcanoyl- or aroyl-groups.

The "precursors" are compounds which are converted into the Component B in the human or animal body. As "active fractions" of the protein the present invention refers to any fragment or precursor of the polypeptidic chain of the compound itself, alone or in combination with related molecules or residues bound to it, for example residues of sugars or phosphates, or aggregates of the polypeptide molecule when such fragments or precursors show the same activity of Component B as medicament.

The present invention refers also to a mixture of polypeptides and derivatives as said above.

A second aspect of the present invention concerns the process of preparation of Component B, such process comprising the isolation of a raw fraction of the protein from a dialysed concentrate of urine after treatment with an adsorbing agent and its purification through

ion exchange chromatography and high resolution chromatography.

Preferably, Component B is prepared through the process illustrated in Figure 1 and comprising the following steps:

a) adsorption of urine at acid pH on kaolin and extraction with ammonia

b)elution of fraction (a) on Bio Rex 70 resin with ammonia

c)elution of fraction (b) on DEAE Sepharose resin with acetate buffer

d) elution of fraction (c) on CM Sepharose resin with acetate buffer

e) elution of fraction (d) on HPLC C18 resin with a mixture of acetate buffer and acetonitrile

f) elution of fraction (e) on DE-52 resin with acetate buffer

g) elution of fraction (f) on D-Zhephyr resin with acetate buffer

h) elution of fraction (g) on HPLC C18 resin with a mixture of aqueous trifluoroacetic acid and acetonitrile

i) elution of fraction (h) on D- Zephyr resin with acetate buffer.

The present invention refers also to recombinant DNA molecules which comprise the nucleotidic sequence encoding the polypeptide according to the invention, its active mutants or fusion proteins, expression vectors which comprise it, host-cells transformed with such vectors and a process of preparation of such polypeptide, its active mutants or fusion proteins, through the culture in appropriate culture media of said transformed cells. The definition "recombinant DNA molecules" include genomic DNA, cDNA, synthetic DNA and combinations thereof. In particular the present invention refers to the nucleotide sequences illustrated in SEQ ID NO: 2 and SEQ ID NO: 3

respectively.

SEQ ID NO: 2: reports the genomic DNA sequence encoding Component B;

Figure 2 reports the restriction map of Component B transcriptional unit;

SEQ ID NO: 3: reports the cDNA sequence encoding Component B;

Figure 8 shows the complete Component B cDNA sequence, in which the restriction sites are indicated. The cloning of Component B can be performed through different techniques. According to one of these techniques an oligonucleotide, or a mixture of oligonucleotides, are prepared, their sequence being derived from the sequence of Component B or its fragment and used as probe for cloning the cDNA or the genomic DNA encoding Component B.

SEQ ID NO: 4: reports the amino acids sequence encoded both by the genomic DNA reported in SEQ ID NO: 2 and by the cDNA reported in SEQ ID NO: 3.

The present invention also refers to recombinant DNA molecules which hybridize with the DNA sequence coding for Component B or fragments thereof.

The gene can contain, or not, the natural introns and can be obtained for example by extraetion from appropriate cells and purification with known methods. Appropriate preparations of DNA, as human genomic DNA, are cut in the appropriate way, preferably with restriction enzymes, and the so obtained fragments are introduced in appropriate recombinant vectors in order to form a DNA library. Such vectors can be selected with synthetic oligonucleotide probes in

order to identify a sequence encoding Component B according to the invention.

In particular, according to the present invention, the genomic DNA of Component B was isolated and cloned.

On the other hand, the corresponding mRNA can be isolated from the cells expressing Component B and used to produce the complementary DNA (cDNA) with known methods. This cDNA after having been converted in the double helix, can be introduced in an apropriated vector which can afterwards be used for transforming an appropriated host cell. The resulting cultures are then selected with an appropriate probe in order to obtain the cDNA encoding the targeted sequences. Once the wanted clone is isolated, the cDNA can be manipulated essentially in the same way as the genomic DNA.

The cDNA does not contain introns.

Because of the degeneration of the genetic code, various codons can be used for encoding a specific amino acid, so that one or more oligonucleotides can be produced, each of them being able to encode fragments of Component B. However only one member of this pool possesses the nucleotide sequence identical to that of the gene. Its presence in the pool and its capacity of hybridizing with the DNA also in the presence of other members of the pool makes it possible the use of the group of non fractioned oligonucleotides in the same way as a single oligonucleotide could be used for cloning the gene encoding the targeted peptide. Alternatively, a single oligonucleotide containing the sequence which is theoretically the most probable being able of encoding the genic fragments of

Component B (according to what described in the "rules for the use of codons" in Lathe R, et al. J.Molec.Biol. 183:1-12 (1985)) allows the identification the complementary DNA encoding Component B or a fragment thereof.

The processes for hybridizing the nucleic acids are known and described, for example in Maniatis T. et al. Molecular Cloning: A laboratory manual, Cold Spring Harbor Press, Cold Spring Harbor, NY, 1982) and in Haymes B.T. et al. Nucleic Acid Hybridization: A practical approach, IRL Press, Oxford, England, (1985). Through the hybridization using said probe or group of nucleotide probes it is possible to identify in a genomic or cDNA gene library the DNA sequences capable of such hybridization which are thereafter analysed to confirm that they encode the polypeptide according to the invention (i.e. Component B). The oligonucleotide which contains such complementary sequence can be synthetized and used as probe to identify and isolate the gene of the polypeptide according to the invention i.e. Component B (Maniatis T. et al. ibid.).

Once the appropriate oligonucleotide specific for Component B is selected using the above said method, it is possible to synthetize and hybridize it with a DNA, or preferably with a cDNA derived from cells capable of expressing the wanted gene preferably after the source of cDNA was enriched of wanted sequences, for example by extraction of the RNA from cells producing high levels of the wanted gene and conversion of the RNA into the corresponding cDNA using the enzyme reverse transcriptase.

Alternatively, the suitable oligonucleotides specific for Component B can be synthesised and used as primers for the amplification of Component B cDNA fragments by RACE-PCR (M. A. Innis et al., PCR Protocols, A Guide to Methods and Applications, Academic Press, 1990).

In particular, according to the present invention, a screening of different human and cellular tissues was performed firstly in order to identify the best available source for mRNA of Component B. Human tissues from brain, kidney, liver, lung, heart, pancreas, placenta, spleen, testis, thymus and uterus as well as epitheliod carcinoma, promyelocytic leukemia, breast adenocarcinoma, Burkitt's lymphoma and myeloma cell lines were screened for this purpose.

The screening was performed by using a sensible assay "reverse transcriptase - polymerase chain reaction" (RT - PCR).

The human uterine tissue resulted the best source of mRNA.

The cDNA clones of Component B were obtained by said tissue using the amplification method named "3' and 5' rapid Amplification of cDNA Ends" (RACE).

The DNA molecules encoding Component B, obtained with the above said method, were introduced in expression vectors constructed with known techniques (Maniatis T et al, ibid.). The double helix cDNA is ligated to plasmid vectors using, for example, techniques comprising the use of synthetic DNA adapters or techniques of binding "blunt-ended".

For the expression of a targeted protein, an expression vector should comprise also specific nucleotide sequences containing the

information regulating transcription and translation bound to the DNA encoding the desired protein in such a way that the expression of the gene and the production of the protein are permitted. First of all, in order that the gene can be transcribed, it must be preceded by a promoter which can be recognised by the RNA polymerase and to which the polymerase binds, thus starting the transcription process.

Many promoters are known operating with different efficency (strong and weak promoters) which are different if used in prokaryotic or eukaryotic cells.

The promoters which can be used in the present invention can be constitutive, as for example promoter int of lambda bacteriophage, promoter Bla of the gene of β-lactamase of pBR322 and the promoter CAT of the gene of chloramphenicol acetyltransferase of pPR325 ecc., or inducible as for example the promoters of prokaryotes as the main right and left promoters of lambda bacteriophage (P1 and Pr), the promoters trp, rec A, lac Z, lac I, ompF and gal of E.coli, or the hybrid promoter trp-lac, etc. (Glik B.R. J. Ind. Microbiol. 1:277-282 (1987).

Together with the strong promoters which are capable of producing huge quantities of mRNA, giving high levels of gene expression in prokaryotic cells, it is necessary to use also binding sites for the ribosomes in order to assure that the mRNA is efficently translated. An example is given by the Shine-Dalgarno (SD) sequence positioned in appropriate way from the starting codon.

For eukaryotic host cells different sequences regulating the transcription and translation can be used according to the nature of the host.

These can be derived from viral sources, as adenovirus, or papilloma virus, Simian virus or similar, wherein the regulation signals are associated to a specific gene having an high level of expression. Possible examples are the promoter TK of the herpes virus, the promoter of SV40, the promoter of gene gal4 of yeast ecc. The signals regulating the starting of transcription can be suitably chosen in order to produce repression or activation in such a way that the expression of the genes can be accordingly modulated.

The DNA molecule comprising the nucleotide sequence encoding Component B of the invention together with the signals regulating transcription and translation are introduced in a vector which is capable of integrating the sequences of the targeted gene in the host cell chromosome.

The cells which bear the introduced DNA in their chromosome can be selected also introducing one or more markers which make it possible to select the host cells containing the expression vector. The marker can provide the cells, for example, with antibiotics resistance or heavy metal (as copper) resistance. The selection gene can be directly bound to the DNA sequences which must be expressed or can be introduced in the cell itself by cotransfection. Other elements may also be necessary for a higher gene expression. These elements can comprise for example transcription enhancers and termination signals and introns. Expression vectors which include

such elements comprise those described by Okayama H. Mol. Cell. Biol. 3: 280 (1983).

Among the factors to be considered for chosing a particular plasmid or viral vector are: the facility of detection of the cells containing the vector which can be easily separated from those which do not contain it; the number of copies of vectors which are wanted in a specific host, and the possibility, or not, of transferring the vector among different host cells.

The preferred prokaryotic vectors comprise plasmids as those capable of replication in E.coli, as pBR322, ColE1, pSC101, pACYC 184 etc. (Maniatis T. et al, ibid.), Bacillus plasmids as pC194, pC221, pT127 etc. (Gryczan T.M. The Molecular Biology of the Bacilli, Academic press, Ny, 307-329 (1982) Streptomyces plasmids as pIJ101 (Kendall K.J. et al. J. bacteriol. 169:4177 - 4183) and Pseudomonas plasmids (John J.F. et al. Rev. Infect. Dis. 8: 693-704 (1986) (Izaki K. Jpn. J. Bacteriol. 33: 729 - 742).

The preferred eukaryotic vectors comprise, for example, BPV, SV40, Baculovirus etc. or their derivatives. Such vectors are known in the art (Bostein D. et al. Miami Wint Symp. 19: 265 - 274) (Broach J.R. The Molecular Biology of the Yeast Saccharomyces: Life Cycle and Inheritance, Cold Spring Harbor, NY, 455-470 (1981) (Broach J.R. cell 28:203-204 (1982) (Bollon D.P. et al. J. Clin. Hematol. Oncol. 10: 39-48 (1980) (Maniatis T. Cell Biology: A Comprehensive Treatise Vol. 3: Gene Expression Acad. Press NY 563 - 608 (1980).

The expression vector so prepared is introduced in the appropriate

host cell with an appropriate method such as transformation, transfection, lipofection, conjugation, protoplastic fusion, electrophoration, precipitation with calcium phosphate, direct microinjection etc. The host cell which can be used for the present invention can be prokaryotic or eukaryotic cells.

Preferred prokaryotes include bacteria as E.Coli, Bacillus, streptomyces, pseudomonas, Salmonella, Serratia, etc.

Particularly preferred is E.coli, as for example strain 294 of E.Coli K12 (AtCC 314446) or E.Coli X1776 (ATCC 31537), E.Coli W 3110 (F, lambda, ATCC 27325).

Preferred eukariotic host cells are mammalian cells as human, monkey, mouse or hamster (Chinese Hamster Ovary, CHO) cells since they assure to the protein molecules post-translation modifications, as for example the correct folding and glycosylation in the right positions.

Yeast cells can be also used for the present invention. There are various recombinant DNA techniques which utilize sequences of strong promoters and a high number of copies of the plasmid and allow the production of the wanted protein in yeast.

After the introduction of the vector in the host cells these are cultivated in a medium which allows the selective growth of cells containing the vector.

The expression of the cloned DNA sequence allows the production of Component B, of a mutant or fragment thereof. The so expressed protein is isolated or purified through conventional techniques comprising extraction, precipitation, chromatography,

electrophoresis, or similar techniques, or affinity chromatography, using anti-Component B antibodies immobilised on the column gel.

Component B can also be produced as milk-secreted protein in transgenic animals.

A further aspect of the present invention is the use of Component B, its salts, functional derivatives, precursors or active fractions as medicament.

In particular Component B has shown anti-inflammatory, anti-coagulant and anti-tumoral properties. Furthermore Component B can be useful in the therapy of pathologies correlated with altered levels of TGF-alpha, such as behavioural and hormonal disturbances, angiogenesys, etc..

In fact, Component B has been shown to inhibit the binding of TGF-alpha to its receptor with an affinity constant is $K_i = 0.77 * 10^{-10}$ M measured by displacement of I^{125} - TGF-alpha from its receptor, obtained from A 431 cell membranes.

The pharmaceutical compositions containing a therapeutically active quantity of Component B in combination with pharmaceutically acceptable excipients or eluents are also an object of the present invention. Such compositions can be formulated for oral, rectal, nasal and particularly parenteral administration.

Also the topic use of Component B is included in the present invention.

The formulations according to the invention include also retard forms as subcutaneous implantations based on liposomes or

microcapsules of copolymers of lactic and glycolic acids.

Other aspects of the invention will be evident in the light of the following detailed description.

Example 1 - Process of preparation of Component B from human urine

The preparation and purification of Component B from human urine is summarized in Fig. 1.

a) STEP 1

The starting material is human urine to which HCl is added up to pH 3.0. After decantation of the precipitate, kaolin is added to the urine (10 g/l of starting urine).

The suspension is left for 16 hours and is thereafter centrifugated. The supernatant is eliminated and kaolin is extracted with ammonia 2M pH 11.0.

The ammonia eluate pH is brought to 8.0 and is concentrated by membrane ultrafiltration (cut off 1000 Daltons). The whole operation is performed at 4°C.

b) STEP 2

The solution of point (a) is added with acetic acid up to pH 4.0 and then with Bio Rex 70 resin previously equilibrated at pH 4.0 with acetic buffer.

The solution is left under stirring for 4 hours and is then filtered on pressfilter.

The adsorbed material is eluted from the Bio Rex 70 resin through elution with ammonia at pH 9.0.

The chromatography eluate is concentrated by membrane ultrafiltration (cut off 1000 Daltons).

The whole operation is performed at 4°C.

c) STEP 3

The material of point (b), equilibrated in acetic buffer pH 5,6 is adsorbed on ion exchange resin like DEAE Sepharose, previously equilibrated at pH 5,6.

At the end of the adsorption elution is performed using ammonium acetate buffer 0.5 M at pH 5.6. The chromatography eluate is concentrated by membrane ultracentrifugation (cut off 1000 Daltons). The whole operation is performed at 4°C.

d) STEP 4

The material of point (c) is equilibrated with acetate buffer at pH 4,5 and adsorbed on ion exchange resin like CM Sepharose previously equilibrated at pH 4.5.

When the adsorption is completed eluation is performed with ammonium acetate buffer 0.15 M pH 4.5. The chromatography eluate is concentrated by membrane ultrafiltration (cut off 1000 daltons). The whole operation is performed at 4°C.

e) STEP 5

The material of step (d) is purified at 25°C by reverse phase chromatography on HPLC C18 resin equilibrated in ammonium acetate buffer 0.05 M pH 5.6.

The adsorbed material is eluted from the resin with an ammonium acetate solution containing acetonitrile 30% (v/v).

The chromatography eluate is concentrate by distillation (40°C) under vacuum.

f) STEP 6

The material of step (e) is purified on ion exchange resin like DE-52, equilibrated at pH 5,6 in ammonium acetate buffer 0,02 M.

The elution of the adsorbed material is performed with buffer 0.25 M. The concentration is performed by membrane ultrafiltration (cut off 1000 Daltons).

The whole operation is performed at 4°C.

g) STEP 7

The material of step (f) is purified on ion exchange resin like D-Zephyr prepacked column (sold by Sepracor), equilibrated at pH 6.2 in 20 mM sodium acetate buffer solution (buffer A).

The elution of the absorbed material is performed by gradient elution from 100% buffer A to 100% 20mM sodium acetate buffer solution at pH 6.2 containing 1M NaCl.

h) STEP 8

The material of step (g) is purified by reversed phase chromatography at 25°C on resin C18 like HPLC.

After adsorbtion the elution is perfomed with linear gradient formed by a binary mixture of aqueous solution of trifluoroacetic acid (TFA 0,1%) and acetonitrile, acidified with TFA (0,1%).

The chromatography eluate concentrated by distillation (45°C) under vacuum and lyophilized.

i) STEP 9

Step 7 is repeated.

The final product, Component B, is recovered as an amorphous white powder.

Example 2 - Analytical characterisation of component B

In order to specify the main physical-chemical characteristics of Component B the purified material from urine underwent the following analytical controls.

a) AMINO ACID SEQUENCE

The amino acid sequence of Component B was determined according to the Edman method

The analysis was performed using a sequencer Applied Biosystem, model 477A, following the indications given by the producer.

Such analysis gave the possibility of identifying the amino acid sequence of Component B relatively to the 81 amino acid residues reported in SEQ ID NO: 1.

b) DETERMINATION OF MOLECULAR WEIGHT

The analysis was performed by "Electron Spray - Mass Spectrometry" (ES-MS) and showed a molecular weight of 8937.9 Daltons. Such analysis has evidenced five disulfide bridges and an 80 Daltons residue attributable to an SO_4 group bound to Tyr (39).

Example 3: Isolation of Human Component B genomic DNA

A human genomic DNA library in lambda phage vector EMBL-3 SP6/T7 was purchased from Clontech (cat. No. HL 1067 J, Lot No. 1221). Genomic DNA was extracted from human placenta and partially digested with Sau 3A. DNA fragments were separated on a sucrose gradient to produce size range between 8 to 22 Kb before cloning into the BamHl site of EMBL-3 Sp6/T7 vector.

Culture media

E. coli K802 cells, purchased from Clontech (cat. No. C1004-1), were cultured in LB medium supplemented with 10 mM $MgSO_4$ and 0.2 % maltose (culture medium).

Phage library was diluted in 0.1 M NaCl, 8 mM $MgSO_4$, 50 mM Tris-Cl pH 7.5, 0.01% gelatin (SM).

The DNA library was plated onto 1.5% agar-LB plates. Top agarose for library plating was: 0.136 M NaCl, 0.6% agarose, 1% tryptone (Merck cat No. 7213).

Hybridization reagents

20xSSC	3 M NaCl, 0.3 M Na citrate, pH 7.0
Hybridization solution	5xSSC, 0.02% SDS, 0.1% N-lauroylsarcosine, 0.5% Blocking reagent (Boehringer cat No. 1096176).
Washing solution A (HRP-oligos) (see below).	3xSSC, 0.1% SDS, urea at various concentrations depending on the specific probe
Washing solution B (^{32}P-oligo CBEX4L)	1xSSC, 0.1% SDS

Detection system

HRP-oligo/DNA hybrids were detected by ECL kit and exposure to Hyperfilm ECL from Amersham (cat. No. RPM 2106 and 2104, respectively).

^{32}P-oligoprobed filters were revealed by exposure to Hyperfilm β-max (Amersham cat. No. RPN10).

Oligonucleotides

Oligonucleotides were synthesized by automatic DNA synthesizer (Applied Biosystem mod. 392).

Oligonucleotides were purified by OPC cartridges (Applied Biosystem cat. No. 400771) or by denaturing PAGE.

Oligonucleotides CB1, CB2 and CBEX2L to be used as probes were 5' modified with N-MMT-C12-aminomodifier (Clontech cat. No. 5206-1) during the last cycle of synthesis.

Horse-radish peroxidase (HRP, Boehringer cat. No. 814393) was conjugated to the modified oligonucleotide according to M. S. Urdea (Nuc. Ac. Res. 16, 4937, 1988), by using 1,4-phenylen-diisothiocyanate (Aldrich cat. No. 25,855-5) as homobifunctional crosslinking agent.

HRP-oligoprobes were purified by anion-exchange HPLC on a Nucleopac PA-100 column (Dionex cat. No. 043010). Elution was performed with 20 mM Na phosphate buffer pH 6.0 and a linear gradient of NaCl from 0.2 to 1.0 M in 30 min..

Purified HRP-oligonucleotides were concentrated by Centricon 10, washed with PBS and stored at 4°C in the dark.

The HRP-oligonucleotide concentration was calculated by OD_{403} (ϵ_{403} =89.5 $cm^{-1}xmM^{-1}$).

The following oligonucleotides were synthesized:

oligo #	sequence	target	
CBPU1	5'TGACTCACACGGCCGGTTCT	promoter	SEQ ID NO: 5
CBPL1	5'CAGCCATGTCCAGTGGTCCT	promoter	SEQ ID NO: 6
CBEX2L	5'ACCACAGCCCATGCTCCA	exon 1	SEQ ID NO: 7

CB1	5'TGCAGGAAGCACTGGTCAT	exon 2	SEQ ID NO: 8
CB2	5'TCTGGCTTGCAGCGGGTAATGGT	exon 2	SEQ ID NO: 9
CB3	5'ATGACCAGTGCTTCCTGC	exon 2	SEQ ID NO: 10
CB5	5'ATCCCACCTGCTGCCTTTTG	intron 2	SEQ ID NO: 11
CBEX4U2	5'CGGCGGCTGGGAGCAGT	intron 2	SEQ ID NO: 12
CBF1	5'ATGGAATTCTAYCCATTYAAYCARTC	exon 3	SEQ ID NO: 13
CBF2	5'ATGGAATTCTAYCCATTYAAYCARAG	exon 3	SEQ ID NO: 14
CBR1	5'GTAGAATTCGCGCCAATGGARTCNGGRTC	exon 3	SEQ ID NO: 15
CBR2	5'GTAGAATTCGCGCCAATGCTRTCNGGRTC	exon 3	SEQ ID NO: 16
CBEX4U	5'AGTACCCCTTCAACCAGAG	exon 3	SEQ ID NO: 17
CBEX4L	5'CAGACACCATGAGTGAGCTG	exon 3	SEQ ID NO: 18
CBEX4U1	5'GACACCTCCTCTGTGACGG	exon 3	SEQ ID NO: 19
CBEX4L1	5'CCAGTTCTGTAGGGTGTCAGT	exon 3	SEQ ID NO: 20

where R = A-G, Y = C-T and N = A-G-C-T.

Library titration

The human genomic DNA library was titred according to standard procedures (F. Ausubel, Current Protocols in Molecular Biology) by infecting 0.3 ml of an overnight culture of E. coli K802 cells with various dilutions of the library, in the range $2x10^{-3}$ to $2x10^{-7}$. Cell-library mixture was incubated at room temperature for 20 min, then transferred to 37°C for 10 min. Infected cells were mixed with 4 ml of top agarose preheated at 50°C and poured onto a 10 cm agar plate prewarmed at 37°C. Plates were incubated at 37°C overnight (ON).

The number of plaques was scored in each plate. Duplicate plates

were prepared for each library dilution. The human genomic DNA library titer was found to be 5×10^9 pfu/ml, as expected.

Library screning

E. coli K802 cells grown overnight at 37°C. 0.6 ml of cell culture were infected with a library aliquot (6×10^4 pfu) suspended in SM. Infection and plating was performed as above, but 9 ml of top agarose and 15 cm plates were used.

Semiconfluent plaques were transferred onto Hybond N^+ nylon membrane (Amersham), according to Amersham instruction manual.

Blotted DNA was denaturated by placing the filters, plaque side up, onto a filter paper soaked in 1.5 M NaCl, 0.5 M NaOH for 7 min.

Blotted DNA was then neutralized by placing the filters onto a filter paper soaked in neutralizing solution (1.5 M NaCl, 0.5 M Tris-Cl pH 7.2, 1 mM EDTA) twice for 3 min each.

Filters were washed in 2xSSC and air-dried. DNA was fixed to the membrane by placing the filters onto a filter paper soaked in 0.4 M NaOH for 20 min.

The filters were finally washed in 5xSSC for 1 min and stored in a plastic bag at 4°C up to hybridization. The human genomic DNA library (1×10^6 clones) was screened at high plating density with HRP-CB2 oligoprobe. 20 positive clones were selected.

Six positive clones were rescreened with both HRP-CB1 and HRP-CB2 oligoprobes. Three clones, named 4D, 12B and 15, were confirmed to be positive for Component B gene.

Hybridization

Filters were preincubated at 42°C for 30 min in the hybridization

solution, then hybridized with the appropriate HRP-oligoprobe (5 ng/ml oligonucleotide moiety in hybridization solution) at 42°C for 45 min and finally washed twice for 15 min each at 42°C in washing solution containing urea at the appropriate concentration (see below).

Filters were washed briefly in 2xSSC at room temperature and the hybridized plaques were detected by ECL reagents and exposure to Hyperfilm for 60 min.

Washing conditions for HRP-oligoprobed filters were experimentally determined to minimize unspecific hybridization to E. coli and lambda phage DNAs. Serial dilutions in the range 500 to 15 attomoles of target DNA were spotted on Hybond N^+ membrane in the presence of lambda DNA (10 ng). Lambda and E. coli DNAs (10 ng each) were used as negative controls. Several strips were prepared and used in hybridization experiments with 5 ng/ml probe. Washings were performed with washing solution A containing 0, 9, 18, 27 and 36% urea.

18% and 27% urea were found effective for CB1 and CB2, respectively. Filters hybridized with CBEX2L were washed with washing solution A containing 18% urea.

Hybridization with ^{32}P-oligo CBEX4L was performed at 50°C and filters were washed at 45°C in washing solution B.

Plaque subscreening

Positive plaques were picked up by a Pasteur pipette and transferred to a tube containing 1 ml of SM plus a drop of

chloroform. After 2 hr incubation under shaking at room temperature, the phage suspension was stored at 4°C.

A 10^{-3} dilution of the phage suspension was plated onto a 10 cm plate and rescreened on two replicate filters with two oligoprobes, i.e. the one used in the first screening and another one matching to an adjacent region of component B.

Independent clones positive with both probes were picked up and resuspended as above.

Preparation of phage stocks

Positive clones were expanded by infecting E. coli K802 cells and growth on 15 cm agar plates. After ON incubation at 37°C, the confluent lysate was collected from agar plates with 10 ml of SM. A few drops of chloroform were added, cell debris were removed by centrifugation at 3000 rpm for 5 min at 4°C and the clear supernatant containing the phages was brought to 50% glycerol, aliquoted and stored at -80°C.

Extraction of phage DNA

$2x10^9$ E. coli K802 cells were infected with the selected phage clone (cell/phage ratio = 4:1) and grown in 100 ml liquid culture medium ON at 37°C. At the end of incubation, full cell lysis was accomplished by adding chloroform (5 µl/ml) to the culture.

Phage DNA was extracted by Quiagen, according to the manufacturer instructions.

Phage DNA sequencing

Phage DNA was sequenced by cycle sequencing kit from Applied Biosystem (cat. No. 401388) with an automated DNA sequencer (Applied

Biosystem mod. 373A). Phage DNA was extracted from clones 4D, 12B and 15 and sequenced by cycle sequencing.

Sequencing primers were derived either from the amino acid sequence of Component B (CBF1, CBF2, CBR1, CBR2) or from the available cDNA or genomic DNA sequencing data.

Sequencing data indicated that the three clones contained the full - length Component B gene.

Phage DNA restriction analysis

Phage DNA was submitted to single and multiple restriction enzyme digestions. DNA fragments were resolved by 0.6% agarose gel electrophoresis and then blotted onto Hybond N^+ membrane. Filters were repeatedly probed with oligonucleotides CBEX2L, CB2 and CBEX4L, matching to exon 1, 2 and 3 respectively.

Subcloning of Component B gene in pBlueScript II SK

Restriction analysis of Component B clone 4D with EcoR1, Xho I and Sfi 1 and Southern blotting with oligoprobes specific for the three exons of Component B indicated that the entire Component B gene was comprised in a 5.2 Kb EcoR1 fragment. (Fig. 5).

Phage DNA was estracted from clone 4D and digested with EcoR1. The resulting DNA fragments were resolved by agarose gel electrophoresis. the 5.2 Kb fragment was purified by Qiaex (qiagen cat. No. 20020) and ligated to EcoR1 linearized pBlueSript II KS (Stratagene cat. No. 212207). E. coli strain XL1-Blue (Stratagene cat. No. 200268) was transformed with the litation mixture and transformed cells were selected on Ap/Tc plates. One clone

containing the expected plasmid, as shown by resctriction analysis with EcoR1, was isolated and named pBSCB4D.

Further restriction analysis was performed with Sma1, Kpn1, Hind III, Sfi1, Acc1, Not1, Sal1, Xho1, EcoRV, Cla1, Hinc II, Hind II, Sca1, Bgl II, Aat 2, Nco1, Nhe1, Hpa1 and Mlu1. In addition Southern blotting was performed with Component B specific oligoprobes on pBSCB4D after single and double digestions.

Fig. 6 shows the restriction map of pBSCB4D plasmid.

Figure 4 shows the restriction map of Component B gene. The Component B gene contains 3 exons separated by 2 introns. The exons are flanked by appropriate consensus acceptor and donor splice sites.

Exon 1 is 84 bp in length and contains 26 nt of untraslated mRNA and the sequence coding for 19 amino acids of a putative signal peptide. It is separated from exon 2 by an intron of 410 bp.

Exon 2 is 120 bp in length and codes for 3 amino acids of a putative signal peptide and 37 amino acids of the mature protein. It is separated from exon 3 by an intron of about 550 bp.

Exon 3 is 326 bp long; it encodes the C-terminal 44 amino acids of Component B and 192 nt of untraslated mRNA, containing a polyadenylation signal (TATAAA) 14 bp upstream to the 3' processing site, to which end the poly(A) tail is attached.

In particular, in the three genomic clones the signal peptide encoding sequence was found to contain a Leu codon at position 11 of the putative signal peptide.

The amino acid sequence of Component B derived from the genomic gene

was found to be identical to the one experimentally determined by Edman degradation.

Sequence analysis upstream to exon 1 evidentiated a promoter region (Figure 3) containing a TATA box at -28 and various upstream promoter elements and enhancers, including a GC-rich box at -58, an AP-1 site at -83, an AP-2 site at -360 and several E boxes.

The TATA box is the preferred binding site for the transcription initiation factor TFIID. The GC-rich box represents the binding site for Sp-1, a general transcription factor involved in the transcritpion of a wide variety of genes (Transcription and Splicing B.D. Hames & D.M. Glover Eds, IRL press, 1988).

The AP-1 site is the binding site for AP-1, the transcription factor complex formed by c-fos and c-jun. The AP-1 site is present in several genes involved in cell growth and differentiation. AP-1 is the one of several cis-elements that mediate induction responses to activators of protein kinase C (The hormonal control of gene transcription P. Cohen & J.G. Foulkes Eds. Elsevier, 1991).

The AP-2 site is the target for AP-2, a transcription factor activated by PMA and cAMP (*ibidem*).

E boxes are common sequences found in several enhancer regions and play an important role in determing the tissue-specific expression of genes. E box contains the sequence CANNTG, with the two internal bases changing according to the specific E box (R. E. Kingston Current Opinion Cell. Biol. 1989; 1, 1081 - 1087).

The Component B promoter contains a potential responsive element for

glucocorticoid receptor (GRE), which indicates that the Component B gene could be induced by glucocorticoids.

Subcloning of Component B gene in a vector for expression in mammalian cells

It is known that the expression of rec-proteins in mammalian cells may be improved by the presence of intron(s). The Component B genomic DNA can be expressed in mammalian cells.

To this end, a 1364 bp fragment spanning Component B gene from +50 to +1413 is excised from pBSCB4D by Pvu II and Nar 1 digestion. Fig. 2 shows the restriction map of Component B transcriptional unit where Pvu II and Nar1 sites are based. The entire Component B gene is reconstituted by ligation of this fragment with a synthetic oligonucleotide reproducing the 5' end of the gene, flanked by a suitable restriction site for the subsequent gene cloning in an eukaryotic expression plasmid.

Example 4: Isolation of Component B cDNA clones

Rapid Amplification of cDNA Ends (RACE), a technique described by Frohman et al., (1988) Proc. Natl. Acad. Sci. USA 85, 8998, was used to obtain partial cDNA clones corresponding to the 5' and 3' ends of the Component B mRNA. The partial clones contained overlapping DNA sequence and thus could be combined to construct the full-length Component B cDNA sequence. A diagram depicting the general strategy used for RACE cloning is shown in Figure 7.

For 3' RACE, the DNA sequence of the second exon of the Component B gene was available and was used to design the gene specific primer CKCB1 (5'-TCAAGTGCTACACCTGCAAGGAG-3') (SEQ ID NO: 21). cDNA

synthesis was primed from the poly A tail of human uterus poly A^+ RNA with the oligonucleotide 5' - GGCCACGCGTCGACTAGTACTTTTTTTTTTTTTTTT-3' (SEQ ID NO: 24), called the adapter primer of AP. The cDNA was used as the template for a polymerase chain reaction (PCR) with the CKCB1 and AP primers, which produced an approximately 450 base pair (bp) fragment corresponding to the 3' end of the Component B cDNA.

For 5' RACE, the primer CKCB7 (5'-CGTCAGAGAGGAGGTG-3') (SEQ ID NO: 22 was designed from the DNA sequence of the 450 bp 3' RACE fragment and was used to prime cDNA synthesis from human uterus poly A^+ RNA. After purification to remove the mRNA and the CKCB7 primer, an oligodeoxycytidine tail was added to the 3' end of the cDNA. The tailed cDNA was used as the template in a PCR with the nested primer CKCB2 (5'-ACCGTCACCAGCGTGGTC-3') (SEQ ID NO: 23) and the anchor primer (ACP, 5'-CTACTACTACTAGGCCACGCGTCGACTAGTACGGGIIGGGIIGGGIIG-3') (SEQ ID NO: 25) or a mixture of the ACP and the universal amplification primer (UAP, 5'CTACTACTACTAGGCCACGCGTCGACTAGTAC-3') (SEQ ID NO: 26). CKCB2 annealed to the 3' end of the second exon sequence and the ACP annealed to the oligodeoxycytidine tail. An approximately 230 bp fragment was obtained which contained DNA sequence corresponding to the 5' end of the Component B mRNA.

General experimental protocols (such as polyacrylamide gel electrophoresis, ethanol precipitation, ligation, and restriction endonuclease digestion), bacterial culture media (such as LB) and chemical solutions (such as phenol) used are described in detail in

Sambrook et al, (1989) Molecular Cloning: A Laboratory Manual, 2nd ed., Cold Spring Harbor Laboratory Press. New York, unless otherwise referenced.

3' RACE cloning procedure

The 3' RACE System for Rapid Amplification of cDNA ends was purchased from Life Technologies, Inc., Grand Island, NY. Human uterus poly A^+ RNA was purchased from Clontech Laboratories, Inc. Palo Alto, CA. First strand cDNA synthesis was done using the protocol and reagents supplied with the 3' RACE system. Briefly, 1 µl (1 µg) of uterus poly A^+ RNA was combined with 1 µl of a 10 µM solution of AP and 12 µl of diethyl pyrocarbonate (DEPC)-treated water and the mixture was heated to 65°C for 10 minutes. After chilling the mixture on ice, the reaction components were added so that the final composition was approximately 20 mM Tris-HCl (pH 8.4), 50 mM KCl, 2.5 mM $MgCl_2$, 100 µg/ml bovine serum albumin, 500 nM AP, 500 µM each of dATP, dCTP, dGTP, and dTTP, and 50 ng/µl RNA in a volume of 19 µl. The reaction mixture was heated to 42°C and 1 µl (20 units) of SuperScript reverse transcriptase was added. After incubation for 30 minutes at 42°C, the reaction mixture was chilled on ice and 1 µl RNaseH (2 units) was added. RNaseH digestion was done for 10 minutes at a temperaure of 42°C. The reaction mixturre was stored at -20°C prior to the PCR.

For PCR, four identical 40 µl mixtures were prepared each with the following composition: 1 µl of uterus poly A^+ cDNA in 40 mM KCl, 70 mM Tris-HCl (pH 8.8), 0.1% Triton X-100, 1 mM $MgCl_2$, 0.25 µM CKCB1, and 0.5 µM AP. Reagents from the 3' RACE system were not used for

PCR. Both the CKCB1 and AP primers were synthesized on an Applied Biosystems, Inc. model 392 oligonucleotide synthesizer. After deprotection, lyophilization, and resuspension in DEPC-treated water the optical density of each solution was measured at a wavelength of 260 nm. Based on the optical density measurements, a 10 µM solution of each oligonucleotide was prepared in DEPC-treated water. The crude oligonucleotide solutions were used for PCR with no further purification. The concentration of crude AP that produced identical results to the AP supplied with the 3'RACE system was experimentally determined; 0.4 µM crude AP was equivalent to 0.2 µM AP from Life Technologies, Inc. in the PCR.

The 40 µl PCR reactions were heated to 94°C in a temperature cycler before adding to each a 10 µl mixture containing the following: 1.25 units AmpliTaq DNA polymerase (Perkin Elmer Cetus, Norwalk, CT), 40 mM KCL, 70 mM Tris-HCl (pH 8.8), 0.1% Triton-X-100, 1 mM $MgCl_2$, and 1 mM each of dATP, dTTP, dGTP, and dCTP. The final concentration of each reagent in the PCR was approximately 1 µl uterus cDNA per 50 µl, 1.25 units AmpliTaq DNA polymerase per 50 µl, 40 mM KCl, 70 mM Tris-HCl (pH 8.8), 0.1% Triton X-100, 1 mM $MgCl_2$, 0.2 µM CKCB1, 0.4 µM AP, and 0.2 mM each of dATP, dTTP, dGTP, and dCTP. After completing a 5 minute incubation at 94°C, a "Touchdown" PCR temperature cycling program was performed according to Don, R.H., Cox P.T., Wainwright, B.J., Baker, K. and Mattick, J.S. (1991) Nucl. Acids Res. 19, 4008, by varying the annealing temperature from 73°C to 63°C.

After PCR amplification, the four reactions were combined and the DNA products were size-fractionated by electrophoresis on a 5% polyacrylamide gel. A DNA product of aproximately 450 bp was excised from the gel and purified by electroelution in dialysis tubing. The eluate was extracted with a 50:50 (v:v) mixture of phenol and chloroform, ethanol precipitated, dried, and resuspended in 10 µl sterile water.

Due to the template independent terminal transferase activity of Taq DNA polymerase and its strong preference for dATP (Clark, J.M. (1988) Nucl. Acids Res. 16, 9677 and Mole, S.E., Iggo, R.D. and Lane D.P. (1989) Nucl. Acids Res. 17, 3319), the purified 450 bp PCR fragment was expected to have a single deoxyadenosine residue at each 3' end. For subcloning and characterizing the PCR fragment, a pBluescriptSK+ (Stratagene, La Jolla, CA) "T-vector" was prepared essentially as described by Marchuk et al. (1991) Nucl. Acids Res. 19, 1154 . The pBluescript plasmid (20 µg) was digested with EcoRV restriction endonuclease and then purified by extraction with a 50:50 (v:v) mixture of phenol and chloroform. After ethanol precipitation, the DNA was treated with 9 units of Taq DNA polymerase for 2 hours at 70°Cin a 50 µl reaction containing 50 mM KCl, 10 mM Tris-HCl (pH 9.0), 0.1% Triton X-100, 1.5 mM $MgCl_2$, and 2 mM dTTP. The vector was again purified by extraction with phenol and chloroform (50:50 v:v) and ethanol precipitation. This procedure resulted in the addition of a single deoxythymidine residue to each 3' terminus, and rendered the vector compatible for insertion of DNA fragments synthetized with Taq DNA polymerase.

The unphosphorylated 450 bp PCR fragment was inserted into the T-vector in a reaction with T4 DNA ligase (New England Biolabs, Beverly, MA) using the conditions recommended by the manufacturer. The ligation reaction was incubated for approximately 72 hours at 16°C and then used to transform competent E. coli XL1-Blue cells (Stratagene, La Jolla, CA). The transformed cells were plated onto LB agar containing 50 µg/ml ampicillin. Prior to plating the cells, 100 µl 2% X-gal (Life Technologies, Inc., Grand Island, NY) and 40 µl 100 mM IPTG (Life Technologies, Inc., Grand Island, NY) were sequentially spread onto the agar surface of each plate and allowed to dry. After an overnight incubation at 37°C, colonies with blue pigment and unpigmented colonies (white) were visible. Plasmid DNA was purified from cultures of 12 white colonies. All 12 isolates contained the 450 bp insert. Five clones were chosen for further analysis: 3CB4, 3CB6, 3CB7, 3CB8 and 3CB9. DNA sequence analysis was done using a Sequenase version 2.0 kit (United States Biochemical, Cleveland, Ohio).

5' RACE cloning procedure

The 5' RACE system for Rapid Amplification of cDNA ends was purchased from Life Technologies, Inc., Grand Island, NY. Human uterus poly A^+ RNA was purchased from Clontech Laboratories, Palo alto, CA. The 5' RACE cloning experiments were done using the protocol and reagents supplied with the 5' RACE system with the following exceptions: (a) the CKCB7, ACP and UAP primers were synthetized on an Applied Biosystem, Inc. model 392 oligonucleotide

synthetizer and prepared as described for 3' RACE cloning, and (b) the "Touchdown" PCR temperature cycling program described for 3' RACE cloning was used for cDNA amplification. First strand cDNA was synthetized as follows: 1 µl (1 µg) of uterus poly A^+ RNA was comnbined with 0.5 µl of a 10 µM solution of CKCB7 and 13.5 µl of DEPC-treated water and the mixture was heated at 70°C for 10 minutes. After chilling the mixture on ice, the reaction components were added so that the final composition was approximately 20 nM Tris-HCl (pH 8.4), 50 mM KCl, 3 mM $MgCl_2$, 10 mM DTT, 200 mM CKCB7, 400µM each of dATP, dCTP, dGTP, and dTTP, and 40 ng/µl RNA in a volume of 24 µl. The reaction mixture was heated to 42°C and 1 µl (220 units) of Superscript II reverse transcriptase was added. After incubation for 30 minutes at 42°C and for 15 minutes at 70°C, the reaction mixture was placed at 55°C and 1 µl RNaseH (2 units) was added. RNaseH digestion was done for 10 minutes at a temperature of 55°C.

The cDNA was separated from unincorporated dNTPs, CKCB7, and proteins unsing a Glassmax DNA Isolation Spin Cartridge (included in the 5' RACE system). Specifically, 120 µl of binding solution (6M NaI) was added to the first strad reaction, and the CDNA/NaI solution was transferred to a GLASSMAX spin cartridge. Following centrifugation at 13,000 X g for 20 seconds, 0.4 ml of cold (4°C) 1X wash buffer was added. The spin cartridge was centrifuged at 13,000 X g for another 20 seconds. This step was repeated two additional times. After washing twice with 400 µl of cold (4°C) 70% ethanol, the cDNA was eluted by adding 50 µl of sterilized, distilled water

to the spin cartridge and centrifuging at 13.000 X g for 20 seconds. A homopolymer tail was added to the 3' end of the cDNA using terminal deoxynucleotidyl transferase (TdT) and dCTP. The tailing reaction was performed in a PCR compatible buffer. 10 µl of purified cDNA was combined with 7.5 µl of DEPC-treated water, 2.5 µl of 10X reaction buffer, 1.5 µl of a 25 mM solution of $MgCl_2$ and 2.5 µl of a 2mM solution of dCTP. The reaction mixutre was incubated for 2 to 3 minutes at 94°C. After chilling for 1 minute on ice, 1 µl of TdT (10 units/µl) was added. The final composition was therefore: 10 µl cDNA in 20 mM Tris-HCl (pH 8.4), 50 mM KCl, 1.5 mM $MgCl_2$, 200 µM dCTP, 0.4 units/µl TdT. The reaction mixture was incubated for 10 minutes at 37°C and then for 10 minutes at 70°C to inactivate the TdT.

For the PCR, four different reactions were prepared with the following final primer concentrations (per 50 µl):

1. 400 nM ACP
2. 800 nM ACP
3. 360 nM UAP and 40 nM ACP (UAP:ACP, 9:1)
4. 720 nM UAP and 80 nM ACP (UAP:ACP, 9:1)

The final concentrations (per 50 µl) of the remaining components in all four reactions were identical: 5 µl of uterus poly A^+ dC-tailed cDNA in 20 mM Tris-HCl (pH 8.4), 50 mM KCl, 1.5 mM $MgCl_2$, 400 nM CKCB2, and 200 µM each of dATP, dCTP, dGTP and dTTP. The components, including the ACP and UAP primers, were mixed in an initial volume of 45 µl and heated to 94°C in a temperature cycler. A 5 µl mixture

of 1.25 units of AmpliTaq DNA polymerase (Perkin Elmer Cetus, Norwalk, CT) in 20 mM Tris-HCl (pH 8.4) was then added to each reaction to bring the final volume to 50 µl. The "Touchdown" PCR temperature cycling program (as described for 3' RACE cloning) was used to amplify 5' cDNA fragment.

After PCR amplification, the four reactions were combined and the DNA products were size-fractionated by electrophoresis on an 8% polyacrylamide gel. A DNA product of approximately 230 bp was excised from the gel, purified, and subcloned into the "T-vector" as described for the 450 bp 3' RACE fragment. Five clones were chosen for further analysis: 5CB2, 5CB3, 5CB5, 5CB6, 5CB11. DNA sequence analysis was done using a Sequenase version 2.0 kit (United States Biochemical, Cleveland, Ohio).

Figure 8 reports the complete Component B cDNA sequence assembled from RACE clones 5CB3 and 3CB7, in which the restriction sites are indicated .

By sequence alignment the cDNA sequence of clones 5cb3, 5cb6, 5cb11 and 3cb4, 3cb7, 3cb9 were shown to perfectly match with the Component B exons (genomic clone 4D of Example 3).

Although the present invention was illustrated with specific examples, it is clear that modifications can be performed on the operations as described always remaining within the spirit and scope of the invention.

LEGENDS TO FIGURES

Figure 1. Flow chart of the process of Component B from urine.

Figure 2. Restriction map of Component B genomic transcriptional

unit (such Component B genomic DNA reported in SEQ ID NO: 2). Arrows indicate the splicing sites.

Figure 3. Sequence of Component B promoter region (said Component B promoter region reported in SEQ ID NO: 2). Binding sites for AP-1, AP-2, Sp-1 and E-boxes transcription factors are indicated. TATA box is indicated. GRE is also indicated.

Figure 4. Restriction map of Component B gene. The derived mRNA is shown below the genomic gene by a line, the boxed region represents the protein encoding sequence.

Figure 5. Restriction map of clone 4D insert.

Figure 6. Restriction map of pBSCB4D plasmid.

Figure 7. General strategy used for RACE cloning of Component B DNA sequence.

Figure 8. Complete Component B cDNA sequence, in which the restriction sites are indicated (said Component B cDNA reported in SEQ ID NO: 3).

CLAIMS

1. Polypeptide comprising the peptide sequence of sequence SEQ ID NO: 1, its salts, functional derivatives, precursors and active fractions or mixtures thereof.

2. Process for the production of the polypeptide according to claim 1 comprising the following steps:

a) adsorption of urine, at pH acid, on kaolin and extraction with ammonia

b) elution of fraction (a) on Bio Rex 70 resin with ammonia

c) elution of fraction (b) on DEAE Sepharose resin with acetate buffer

d) elution of fraction (c) on CM Sepharose resin with acetate buffer

e)elution of fraction (d) on HPLC C18 resin with a mixture of acetate buffer and acetonitrile

f) elution of fraction (e) on DE-52 resin with acetate buffer

g) elution of fraction (f) on D-Zephyr resin with acetate buffer

h) elution of fraction (g) on HPLC C18 resin with a mixture of aqueous trifluoroacetic acid and acetonitrile

i) elution of fraction (h) on D-Zephyr resin with acetate buffer.

3. Process according to claim 2 wherein the urine is human urine.

4. DNA molecule comprising the DNA sequence coding for the polypeptide according to claim 1, its mutant or active fraction.

5. DNA molecule which hybridizes with the DNA molecule according to claim 4 and codes for a polypeptide according to claim 1, its mutant or active fraction.

6. Genomic DNA molecule comprising the nucleotide sequence of SEQ ID NO: 2.

7. cDNA molecule comprising the nucleotide sequence of SEQ ID NO: 3.

8. Expression vector comprising the DNA molecule according to any of claims 4 to 7.

9. Host cell transformed with an expression vector according to claim 8.

10. Process for the production of the polypeptide according to claim 1 comprising the culture of a host cell transformed according to claim 9 and the recovering of the protein from said cell or from the culture medium.

11. Essentially pure protein, its salts, functional derivatives, prodrug and active fractions, obtainable through a process comprising the following steps:

a) adsorption of urine, at pH acid, on kaolin and extraction with ammonia

b)elution of fraction (a) on Bio Rex 70 resin with ammonia

c)elution of fraction (b) on DEAE Sepharose resin with acetate buffer

d) elution of fraction (c) on CM Sepharose resin with acetate buffer

e)elution of fraction (d) on HPLC C18 resin with a mixture of acetate buffer and acetonitrile

f) elution of fraction (e) on DE-52 resin with acetate buffer

g) elution of fraction (f) on D-Zephyr resin with acetate buffer

h) elution of fraction (g) on HPLC C18 resin with a mixture of

aqueous trifluoroacetic acid and acetonitrile

i) elution of fraction (h) on D-Zephyr resin with acetate buffer.

12. Polypeptide according to claim 1 for use as a medicament.

13. Use of the polypeptide according to claim 1, for the preparation of a medicament having anti-inflammatory and/or anti-coagulant and/or anti-tumoral activity.

14. Pharmaceutical composition comprising a therapeutically acceptable quantity of a polypeptide according to claim 1, or its salt, functional derivative, precursor, active fraction or mixture thereof in combination with one or more therapeutically acceptable eccipient or eluent.

15. Protein according to claim 11 for use as medicament.

16. Pharmaceutical compositions comprising a therapeutically acceptable quantity of a protein according to claim 11, or its salts, functional derivatives, precursors, active fractions or mixture thereof in combination with one or more pharmaceutically acceptable eccipients or eluents.

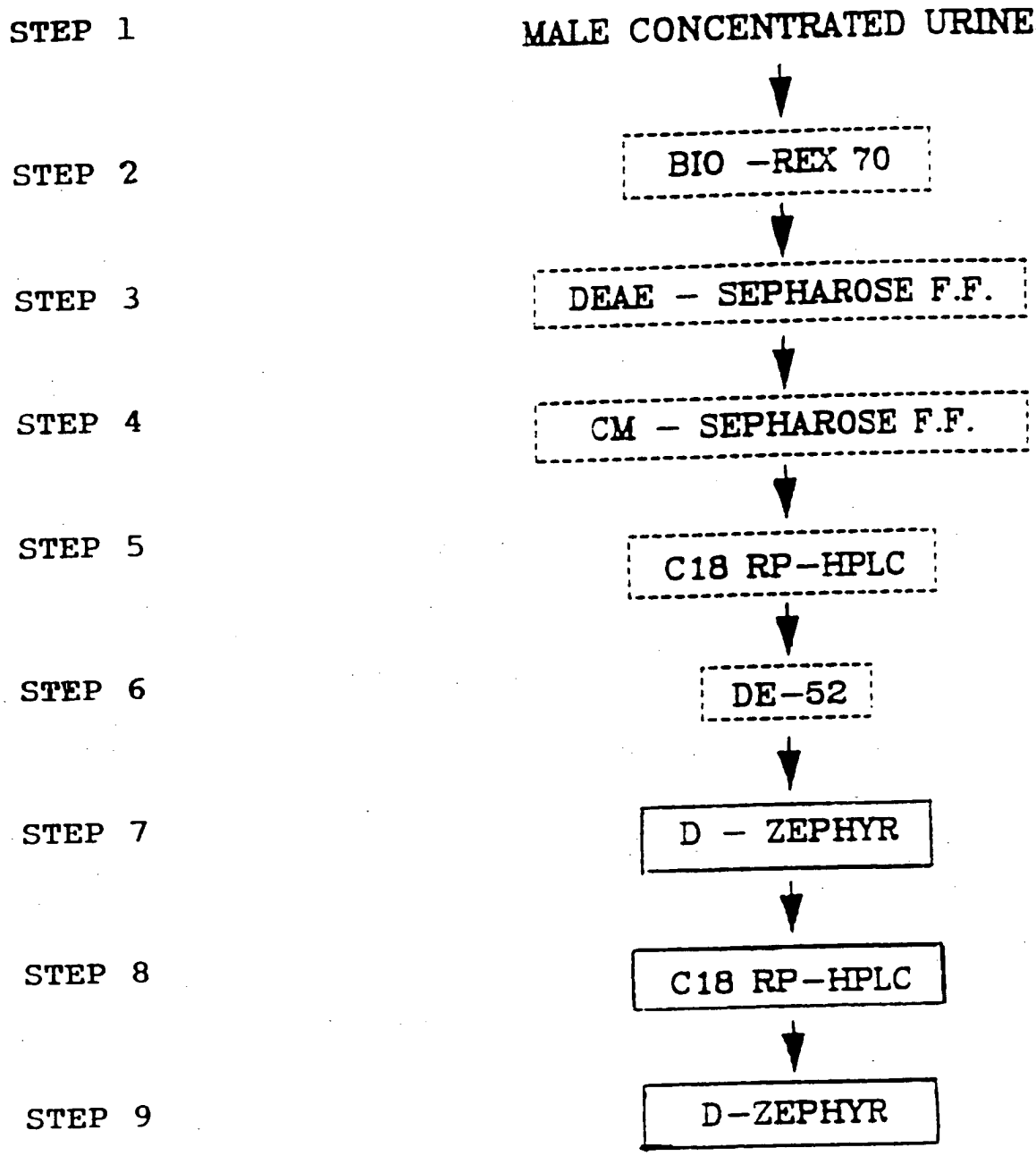
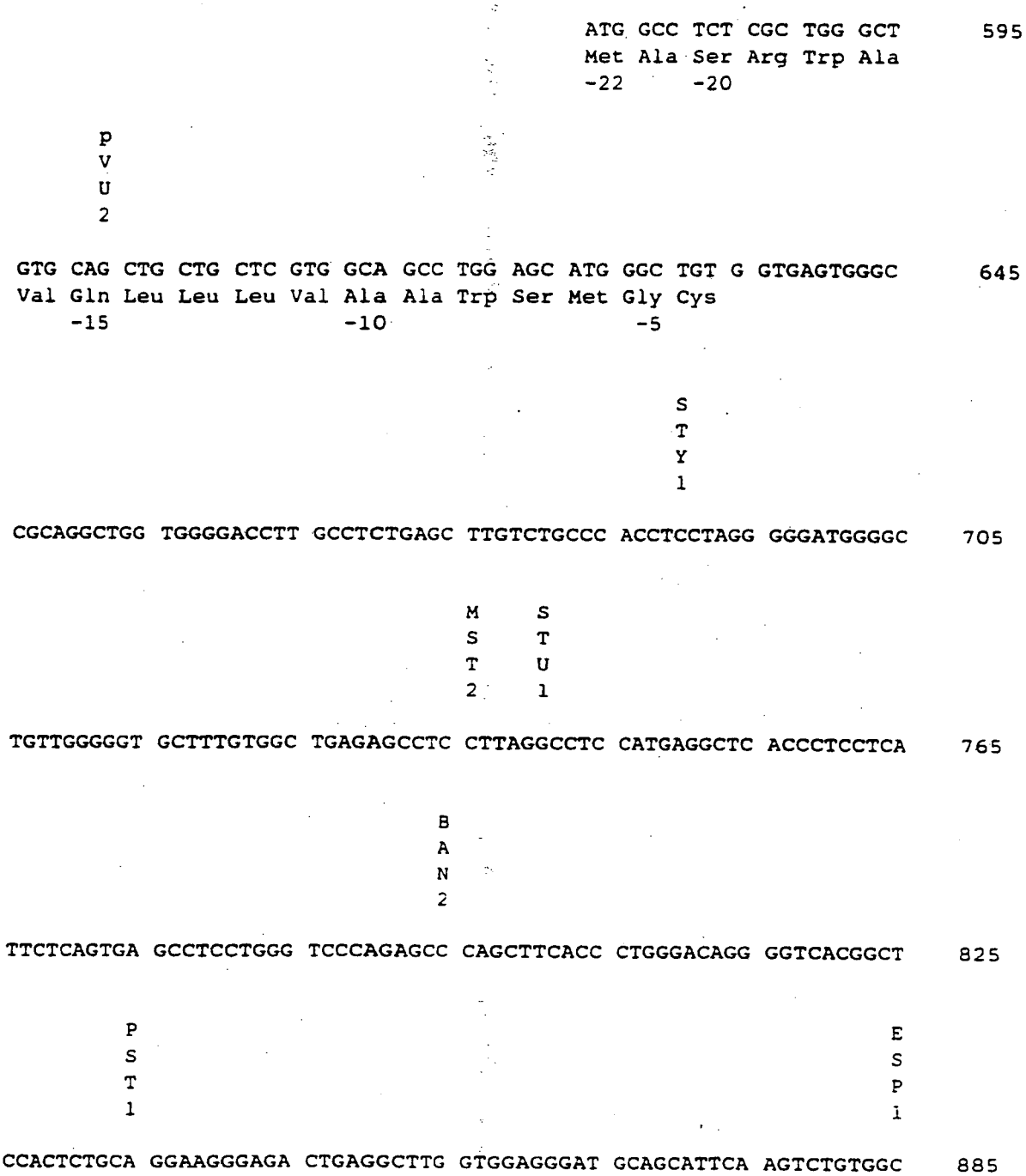


Fig.1

Flow chart of the purification of Component B

FIGURE 2: Restriction map of Component B genomic transcriptional unit



```
                                                    A S
                                                    V M
                                                    A A
                                                    1 1

TCAGCTCAGT TAGAGAAAGC TGCCAGAGAG GCCCCTTGAA GGSCTGCCCG GGGCCTTGAA     945

AGATGTCAGC GAGACTCCTT CAGCCCCTGC CTCCTGGTTC CAGGATGAGV CCACCGAGGT    1005

CAGGTGATGA GGTTCTGCCC CCATCCCTCA CCCAG GT GAG GCC CTC AAG TGC         1057
                                        Gly Glu Ala Leu Lys Cys
                                         -3          1

                          B                         P
                          A                         S
                          N                         T
                          2                         1

TAC ACC TGC AAG GAG CCC ATG ACC AGT GCT TCC TGC AGG ACC ATT ACC      1105
Tyr Thr Cys Lys Glu Pro Met Thr Ser Ala Ser Cys Arg Thr Ile Thr
      5                  10                  15

CGC TGC AAG CCA GAG GAC ACA GCC TGC ATG ACC ACG CTG GTG ACG GTG      1153
Arg Cys Lys Pro Glu Asp Thr Ala Cys Met Thr Thr Leu Val Thr Val
 20                  25                  30                  35

GAG GCA  G GTGAGGCCAG GCCCCACGGC AGCCCTGGGT GCAGTGGAGT CAGGGCCACC    1210
Glu Ala

TCCCCCAAGT GCGTCCCTCC TTTGCTGGTG CTCCTCCCGG CCCAAAAGGA AGCAGGTGGG    1270

                          S
                          T
                          Y
                          1

ATGGGCAGAA CAGGCTGCCA CACCTTGGCA GGGGTGCCTT CCACGAGGGT GGCACAGCCC    1330

CCTCAGAGAC CCAGTCCTGG GGCACCAGGC GCTGGAGGTG GGTGGGGCTT AATGGCCGGG    1390

   K          B
   P          A
   N          N
   1          2

GTACCCTGGG GGGCTCAAAC CCCAGCTCTG ACACAGACCC ACTGGGTGGT GTTGCCACAG    1450
```

```
          AB      B                                                S
          VA      A                                                T
          AN      N                                                U
          12      2                                                1

CCTCTGGGCT CGGGCTCCCA TCTCAGCGCA GGCACTTCAG AGGTCTGACA AGGCCTAATA    1509

                                        B
                                        A
                                        N
                                        2

ATTCATGAAC AGGTCACAGT CAGAGGAGGG CTGGGCCCTG GGTGGCTTCA CAGATGTGGA    1570

CTATTGGGAA CAGGGATCAC AGGGAGGKTG AGGTCAGSCG ACGGCGGCTG GGAGCAGTGC    1630

                    P                                          P
                    S                                          S
                    T                                          T
                    1                                          1

AGCAGCAGGC AGGCGCTGCA GGGGAGTGAG GGTTCTGACA CTGGCCCACC CTGCAG  AG    1688
                                                               Glu

                          B     B
                          A     S
                          N     T
                          2     E

TAC CCC TTC AAC CAG AGC CCC GTG GTG ACC CGC TCC TGC TCC AGC TCC      1736
Tyr Pro Phe Asn Gln Ser Pro Val Val Thr Arg Ser Cys Ser Ser Ser
     40                  45                  50

        B
        A
        L
        1

TGT GTG GCC ACC GAC CCC GAC AGC ATC GGG GCC GCC CAC CTG ATC TTC      1784
Cys Val Ala Thr Asp Pro Asp Ser Ile Gly Ala Ala His Leu Ile Phe
 55                  60                  65                  70
```

```
TGC TGC TTC CGA GAC CTC TGC AAC TCG GAA CTC TGAACCCAGG GCGGCAGGGC    1837
Cys Cys Phe Arg Asp Leu Cys Asn Ser Glu Leu
                75                  80

            M
            S
            T
            2

GGAAGGTGCT CCTCAGGCAC CTCCTCTCTG ACGGGGCCTG GCTCCACCTG TGATCACCTC    1897

CCCCTGCTTC CTGCTGCTGT GGCACAGCTC ACTCATGGGG TCTGAGGGGA GAGAAGCACA    1957

      N
      A
      R
      1

CCAGGGGCGC CCTCTGCCTT CCATACCCCA CGCTTATAAA ACATAACTAA GCCAAGCGTG    2017

GACATGACTT TTGT                                                      2031
```

FIGURE 3: Sequence of Component B promoter region

```
                    E box   E box                                  E box
TGGCCCATGC TACCCTCACC TGACACCTGC TTCCTACCTC TGGTTTCTAC TTTGCAGGTG      60

      E box
TGTATCAGGT GTACACAGAC CAGGTAGAGG TCTGTGGAGA GGGCTGCAGG CCAGGCTGCA      120

GGGAAGGGGT GCCAGGCGGG GCTAGAGCAA CAAGGGCAGA GGCTACACTG AACCTGGGTC      180

              AP-2
YTAAGGGTCC CCCAGGCTGG GGCTGGGTGG CCTATGTGAA CCCCAGAGGC ACAGCCAGGA      240

CATGGGGGCT CATCAGAGGG GCAGTCTGAG CTCAGCAGGA AAGGCCTTCT CTGTCAGAGC      300

                                   GRE
TGTCCCAGGA CCACTGGACA TGGCTGAGGA ACAGTGAGTT CCCCAGTGTT GGAGGTGTGC      360

AAGCAGAGGC CTGGCCATCG TCCTCAGACA CAGCTCCCAG ATCCAGCTCC CTGCCCGTCT      420

                E box                                   AP-1
GCCATGTTCC TGCCAGCTGC CTCCCCACTG GGCCCTTTAC CACGTTCCTG ACTCACACGG      480

                Sp-1                          TATA box
CCGGTTCTGC CACCGCCCAG AAGCCGGTGC CCAAGGGCCT GGCTATAAAT CCTTGATGTG      540

AGGCTGGCTA CCTCTCATCA CTTCTGAGCA CGGAGCA
```

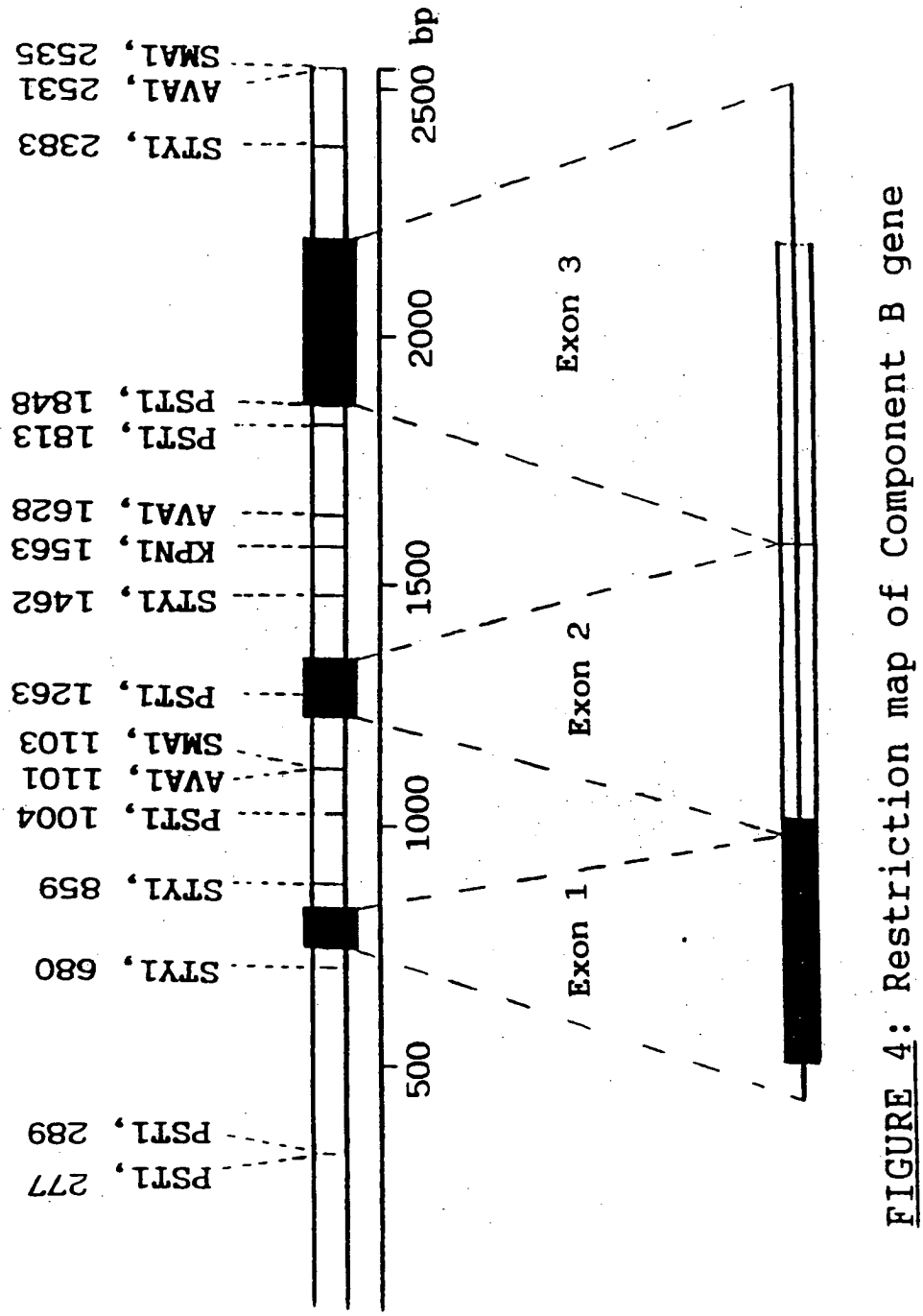


FIGURE 4: Restriction map of Component B gene

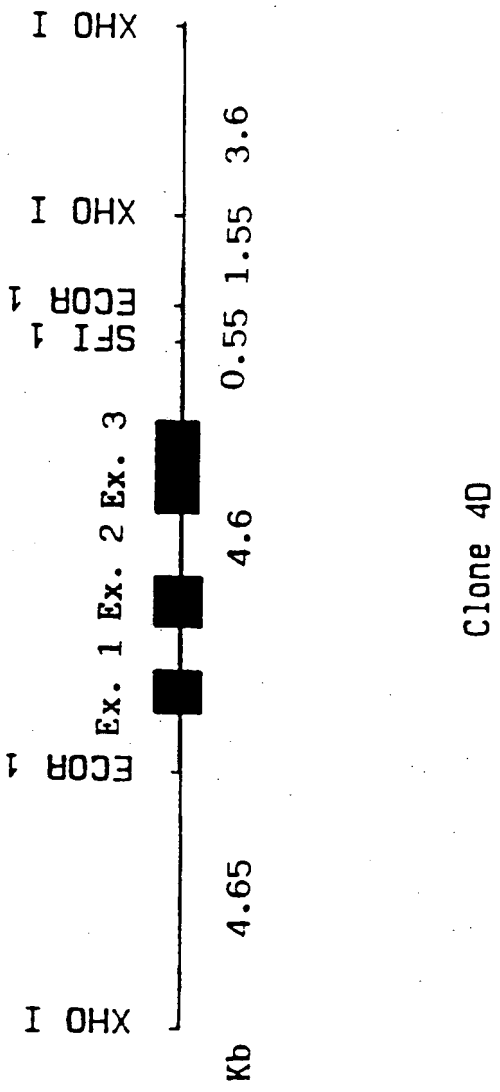


FIGURE 5: Restriction map of clone 4D insert

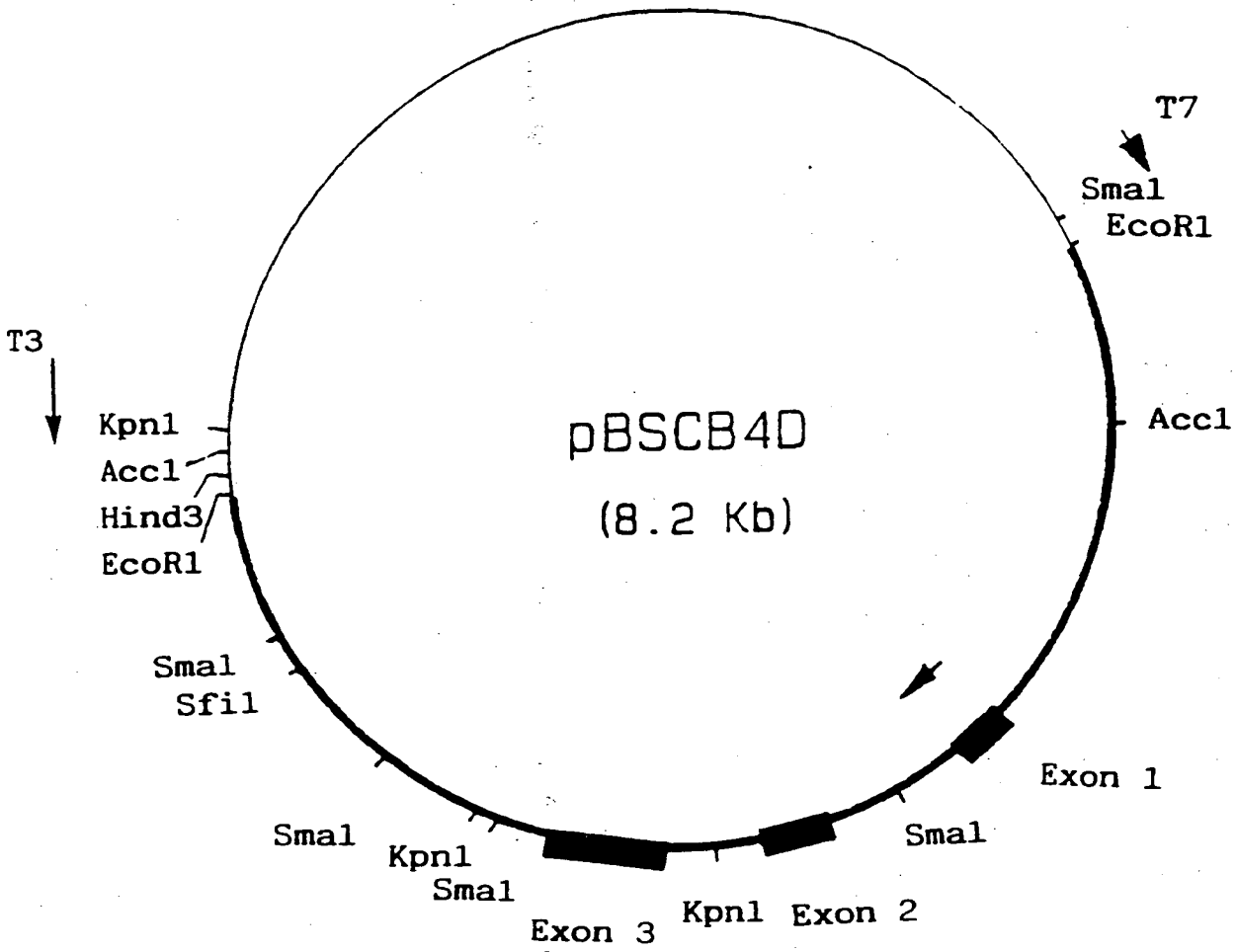
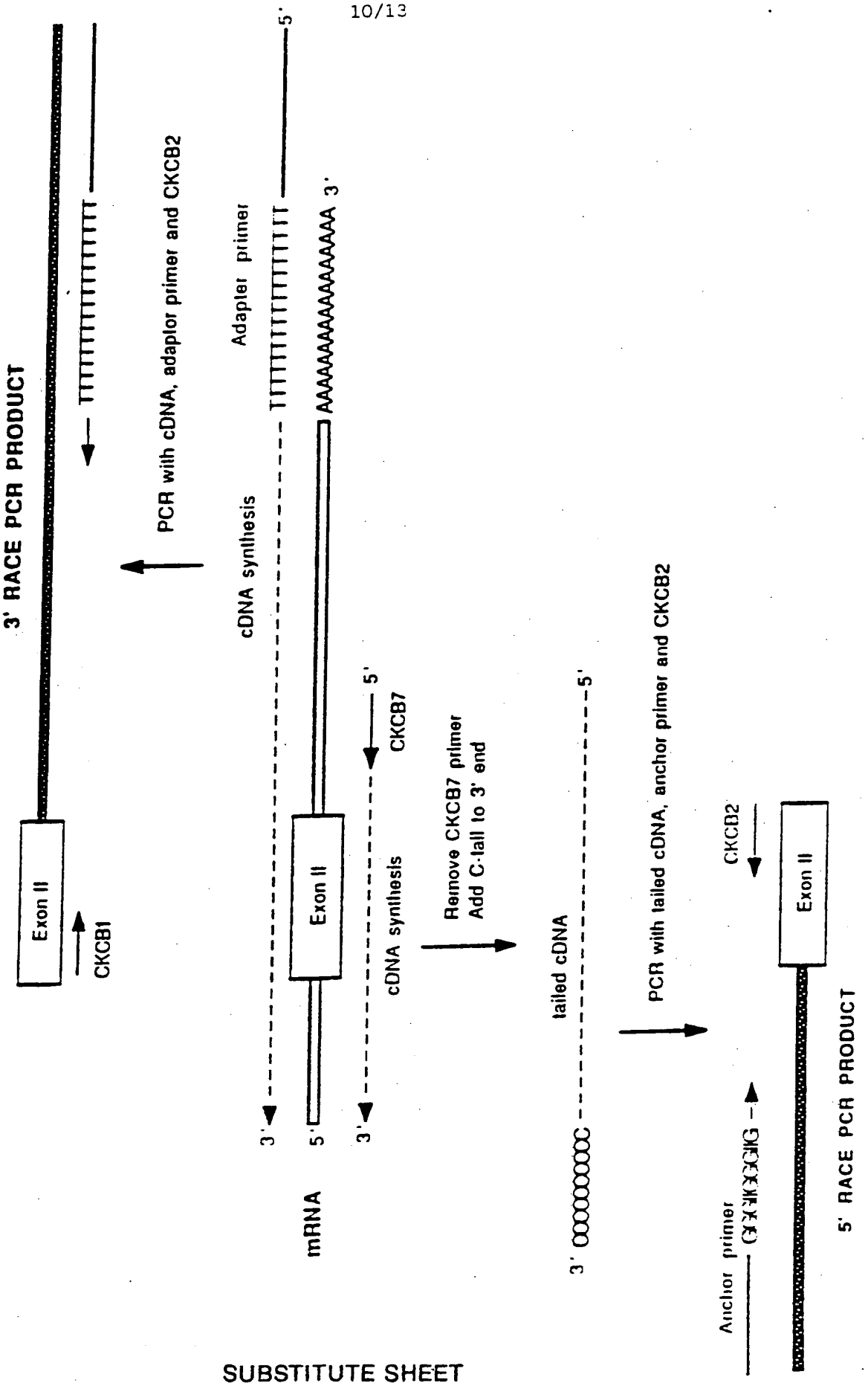


FIg. 6 Restriction map of pBSCB4D

FIGURE 7



SUBSTITUTE SHEET

FIGURE 8 : The complete Component B cDNA sequence assembled from Race clones 5CB3 and 3CB7

```
                 B
                 s
                 As
                 11
                 w2
                 28
                 16
                 II
                  /
                                 B                                  N
                                 Hs                                 sP  AA
                                 ar                                 pv  1p
                                 eD                                 Bu  wa
                                 II                                 II  NB
                                                                     II  II
                                                                      /
ATCACTTCTG AGCACGGAGC A ATG GCC TCT CGC TGG GCT GTG CAG CTG CTG                51
                        Met Ala Ser Arg Trp Ala Val Gln Leu Leu
                        -22     -20                 -15

                                                E
                                                c
                                                o
                                 B              O
                                 c              B1
                 B               e              p0
                 s               8              m9
                 g               3              II
                 I               I               /
CTC GTG GCA GCC TGG AGC ATG GGC TGT GGT GAG GCC CTC AAG TGC TAC                99
Leu Val Ala Ala Trp Ser Met Gly Cys Gly Glu Ala Leu Lys Cys Tyr
       -10                  -5                      1

                      B                                 S
                      s                                 s
                      p                                 e                   N
                 B    B1                                8                   s
                 s    a2                           S    P3                  p
                 p    n8                           f    s8                  B
                 M    I6                           c    t7                  I
                 I    II                           I    II                  I
                       /                                 /
ACC TGC AAG GAG CCC ATG ACC AGT GCT TCC TGC AGG ACC ATT ACC CGC               147
Thr Cys Lys Glu Pro Met Thr Ser Ala Ser Cys Arg Thr Ile Thr Arg
  5                 10                  15                  20

                                                M
                                                s
                                                1
                                                I
TGC AAG CCA GAG GAC ACA GCC TGC ATG ACC ACG CTG GTG ACG GTG GAG               195
Cys Lys Pro Glu Asp Thr Ala Cys Met Thr Thr Leu Val Thr Val Glu
             25                  30                  35
```

```
                                    B
                                    s
                                    p      B
                                    B1     s          B
                                    a2D    Bt         s
                                    n8s    pE         r
                                    I6a    mI         B
                                    III    II         I
                                     //     /
GCA GAG TAC CCC TTC AAC CAG AGC CCC GTG GTG ACC CGC TCC TGC TCC        243
Ala Glu Tyr Pro Phe Asn Gln Ser Pro Val Val Thr Arg Ser Cys Ser
            40                   45                   50

     A
     lX       E  HM
     wc       a  as
     Nm       e  ec
     II       I  II
                 /
AGC TCC TGT GTG GCC ACC GAC CCC GAC AGC ATC GGG GCC GCC CAC CTG        291
Ser Ser Cys Val Ala Thr Asp Pro Asp Ser Ile Gly Ala Ala His Leu
        55                  60                  65

     B        B                                  B
     c        s                                  c
     g        a                                  g
     I        I                                  I
ATC TTC TGC TGC TTC CGA GAC CTC TGC AAC TCG GAA CTC TGAACCCAGG        340
Ile Phe Cys Cys Phe Arg Asp Leu Cys Asn Ser Glu Leu
     70                  75                  80

                       B                       E
                       s                       c
                    A  p  B                    o
         B          1  1Bs                     0
         sE         w  2su  B                  1
         ec         2  8e3  a                  0
         Ri         1  6R6  n                  9
         II         I  III  I                  I
          /            / /
GCGGCAGGGC GGAAGGTGCT CCTCAGGCAC CTCCTCTCTG ACGGGGCCTG GCTCCACCTG     400

B
c
l
I
TGATCACCTC CCCCTGCTTC CTGCTGCTGT GGCACAGCTC ACTCATGGGG TCTGAGGGGA     460
```

```
                  B   H
                  BsN a
                  aaa e
                  nHr I
                  III I
                    /
GAGAAGCACA CCAGGGGCGC CCTCTGCCTT CCATACCCCA CGCTTATAAA ACATAACTAA      520

                                                                       540
GCCAAAAAAA AAAAAAAAAA
```

INTERNATIONAL SEARCH REPORT

Inter nal Application No
PCT/EP 93/03645

A. CLASSIFICATION OF SUBJECT MATTER
IPC 5 C12N15/12 C07K7/10 C12N5/10 A61K37/02

According to International Patent Classification (IPC) or to both national classification and IPC

B. FIELDS SEARCHED

Minimum documentation searched (classification system followed by classification symbols)
IPC 5 C07K A61K

Documentation searched other than minimum documentation to the extent that such documents are included in the fields searched

Electronic data base consulted during the international search (name of data base and, where practical, search terms used)

C. DOCUMENTS CONSIDERED TO BE RELEVANT

Category *	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
A	CLINICA CHIMICA ACTA vol. 207 , 1992 pages 239 - 249 A. BERNARD ET AL 'Human urinary protein 1: Evidence for identity with the Clara cell protein and occurence in respiratory tract and urogenital secretions'	
A	WO,A,82 04261 (UNIVERSITY OF TENNESSEE RESEARCH CORPORATION) 9 December 1982	
A	EP,A,0 069 232 (BURZYNSKI, S.R.) 12 January 1983	

☐ Further documents are listed in the continuation of box C.

☒ Patent family members are listed in annex.

* Special categories of cited documents :

"A" document defining the general state of the art which is not considered to be of particular relevance

"E" earlier document but published on or after the international filing date

"L" document which may throw doubts on priority claim(s) or which is cited to establish the publication date of another citation or other special reason (as specified)

"O" document referring to an oral disclosure, use, exhibition or other means

"P" document published prior to the international filing date but later than the priority date claimed

"T" later document published after the international filing date or priority date and not in conflict with the application but cited to understand the principle or theory underlying the invention

"X" document of particular relevance; the claimed invention cannot be considered novel or cannot be considered to involve an inventive step when the document is taken alone

"Y" document of particular relevance; the claimed invention cannot be considered to involve an inventive step when the document is combined with one or more other such documents, such combination being obvious to a person skilled in the art.

"&" document member of the same patent family

Date of the actual completion of the international search	Date of mailing of the international search report
19 April 1994	20. 05. 94

Name and mailing address of the ISA	Authorized officer
European Patent Office, P.B. 5818 Patentlaan 2 NL - 2280 HV Rijswijk Tel. (+ 31-70) 340-2040, Tx. 31 651 epo nl, Fax (+ 31-70) 340-3016	Van der Schaal, C

Form PCT/ISA/210 (second sheet) (July 1992)

INTERNATIONAL SEARCH REPORT

Information on patent family members

Inter nal Application No
PCT/EP 93/03645

Patent document cited in search report	Publication date	Patent family member(s)		Publication date
WO-A-8204261	09-12-82	US-A-	4359415	16-11-82
		BE-A-	893393	01-10-82
		EP-A,B	0080502	08-06-83
		JP-A-	58000991	06-01-83
EP-A-0069232	12-01-83	US-A-	4470970	11-09-84
		AU-B-	551109	17-04-86
		AU-A-	8423982	06-01-83
		CA-A,C	1188218	04-06-85
		JP-A-	58010521	21-01-83
		JP-A-	5058886	09-03-93
		JP-A-	5032548	09-02-93
		US-A-	4558057	10-12-85
		US-A-	4559325	17-12-85